Towards efficient search tools for biomedical databases:

Characterizing user search habits and recognizing their information needs

Rezarta Islamaj Doğan, G. Craig Murray, Aurélie Névéol and Zhiyong Lu

National Center for Biotechnology Information, National Library of Medicine, Bethesda, MD 20894.

Overview and data analysis

Efficient search tools are crucial for researchers to identify literature concerning their own research. Finding citations relevant to a user's information need is tightly coupled with *understanding* the user's needs and search behavior.

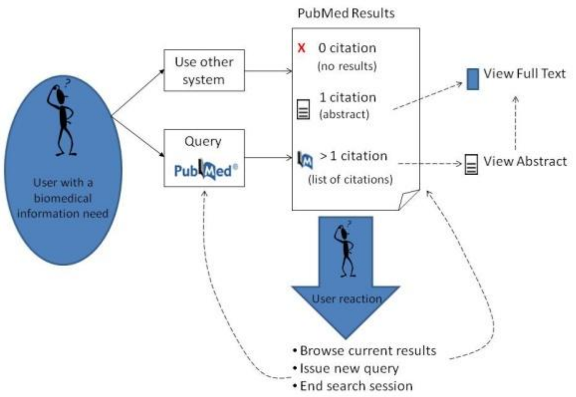

Fig.1. An overview of user interactions with PubMed

We analyzed a collection of PubMed® logs that contained 100 million user queries, abstract views and full text views.

Of these, 10,000 user queries were manually reviewed and categorized into search requests categories.

PubMed users' information needs

Biomedical information queries are short. Each word has significant impact for results. Study of user queries helps with *understanding of* users' needs and efficiently directing them to the useful articles.

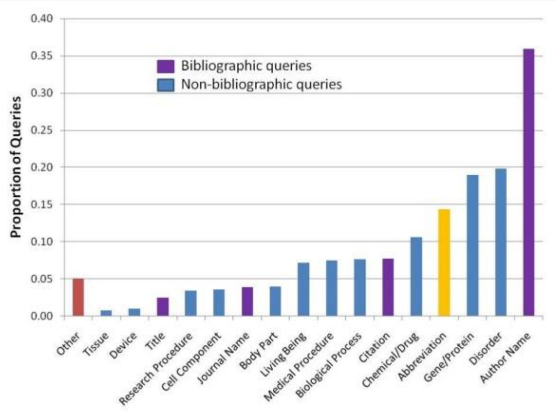

Fig.2. A summary of user query categories in PubMed

The most popular types of search are:

Author Name	36 %
Disease Name	20 %
Gene/Protein name	19 %

Gene/Protein references are often abbreviated Author Name queries frequently include other Citation information.

PubMed users' search habits

PubMed users are persistent in seeking information; they reformulate their queries often and click on 4 citations per query (avg). Their decisions to click on articles are influenced by the result set size. The top ranked citation is clicked 28% of the time.

Results of query analysis	
Average number of queries issued by a user per day	4.05
Average number of words in a PubMed query	3.54
Average number of citations returned per query	44
Results of click through analysis	
Queries that do not retrieve any results	15 %
Queries that were followed by another query	47 %
Abstract views followed by full text of the same article	29 %
Average number of abstract or full text articles requested (clicked) by a user per day	3.57

Table 1. Highlights of PubMed users search behavior

Our analysis can be used to improve retrieval quality and inform future development for PubMed and other biomedical search engines.

Our findings suggest that specialized techniques might be more desirable than traditional information retrieval techniques.

ACKNOWLEDGEMENTS: This research was supported by the Intramural Research Program of the NIH, National Library of Medicine.

REFERENCES: *Islamaj Dogan R, et al. (2009) Understanding PubMed® user search behavior through log analysis*. Database. Vol. 2009: bap018;